2025中国金融行业大模型产业洞察

金融智慧升级，大模型赋能未来

企业标签：阿里云、华为云、商汤科技

AI变革行业创新发展

China Financial Large Model Industry

中国金融モデル産業

头豹研究院

研究框架

名词解释

- **金融大模型：**指应用于金融领域的大型语言模型，它拥有大量参数和复杂结构，通常基于机器学习和人工智能技术。这些模型通过分析金融相关数据，并基于历史数据和主流的金融理论进行训练，从而能够识别和预测市场趋势，制定相关策略，提高金融决策的精度和效率。
- **金融级AI原生：**描述专为满足金融行业最严格需求而设计和优化的AI系统，包含安全性、可靠性、可扩展性、合规性等方面，适用于高复杂性和高风险的金融场景。
- **生成式AI：**一种人工智能模型，关注学习输入数据的分布规律，并生成与输入数据类似的新数据，广泛应用于图像生成、文本生成等领域。
- **判别式AI：**一种通过分析输入和输出之间的关系来进行分类或回归的AI模型，常用于推荐系统、风险控制等任务。
- **RAG（检索增强生成技术）：**结合信息检索与生成技术的系统，用于应对复杂查询和生成任务，如问答系统和内容创作。
- **大模型规模定律（Scaling Law）：**描述大模型参数数量与性能提升的关系，强调模型规模扩大会显著提升其学习能力和任务处理性能。
- **低延时与高并发：**大模型处理实时任务（如交易监控）和高并发场景（如大规模用户同时请求）的关键能力，依赖模型优化和分布式架构实现。
- **模型微调：**在预训练模型的基础上，利用特定领域的数据对模型进行调整，以提升其在特定任务中的表现。
- **云原生架构：**通过容器化、微服务等技术实现资源弹性和高效管理，为大模型训练和推理提供灵活的基础设施。
- **金融云专属VPC模式：**一种部署形式，将大模型应用和知识库部署在金融云客户的专属虚拟私有云中，确保数据隐私和安全。

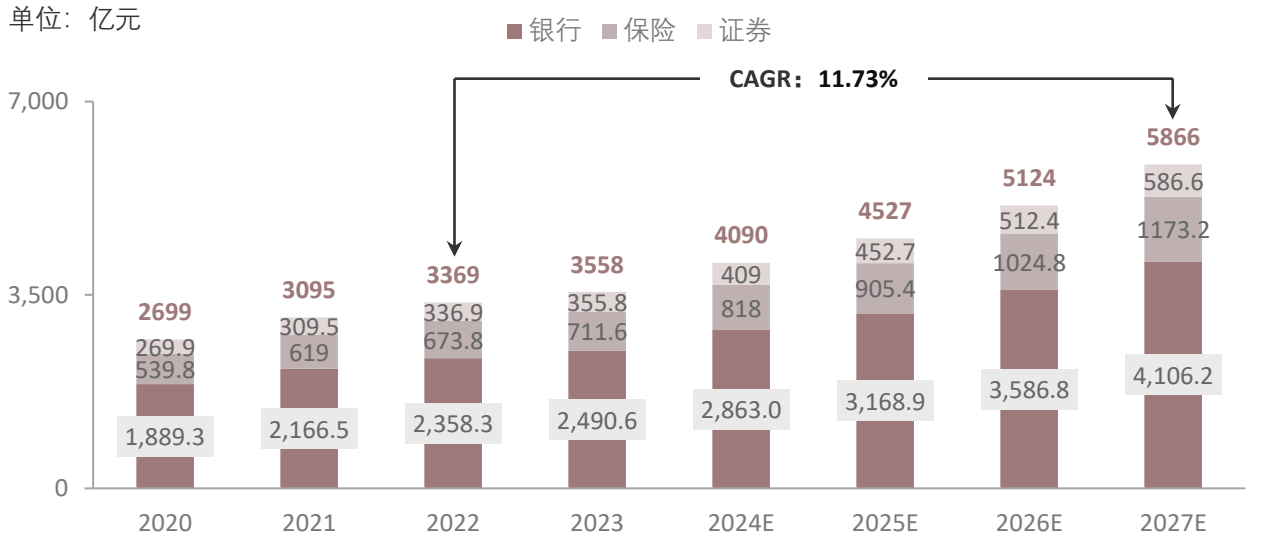
Chapter 1
金融大模型产业洞察

- ❑ 发展背景（产业）
- ❑ 发展背景（技术）
- ❑ 发展背景（转型进度）
- ❑ 开发框架
- ❑ 产业政策
- ❑ 发展趋势
- ❑ 落地挑战
- ❑ 业务场景
- ❑ 市场规模

中国金融大模型产业洞察——发展背景（产业）

- 持续增长的金融科技投入和核心技术创新，为金融大模型提供了数据、算力和场景三大核心支撑，加速其在智能风控、精准营销、自动决策等业务领域的深度应用和价值创造

中国金融机构科技投入情况，2020-2027年



- 从2022年到2027年，中国金融机构在科技投入方面保持**年复合增长率（CAGR）为11.73%的稳定增长，预计总投入将从2022年的3,369亿元增至2027年的5,866亿元。其中，银行业（70%）始终占据主导地位，而保险业（20%）**和**证券业（10%）**的投入占比逐年提升。这一持续增长的科技投入，尤其在数据治理、计算力建设、智能服务等方面的加码，为**金融大模型的研发、部署和落地**提供了**优渥的土壤**

银行	2023年金融科技投入（亿元）	同比增长（%）	占营收比重（%）	主要举措
工商银行	272.46	3.9	3.23	升级ECOS技术生态，部署云和分布式技术体系，建成千亿级AI大模型技术体系，深化D-ICBC数字生态
建设银行	250.24	7.44	3.25	推进核心技术自主可控能力建设，构建“建行云”基础设施，推动核心业务系统分布式转型，自研人工智能技术平台
农业银行	248.5	7.06	3.58	加快新一代技术体系转型，打造数字基础设施，推进云原生能力建设，探索区块链和量子技术
中国银行	223.97	3.97	3.59	加快数字化创新，建设“多地多中心”基础设施，推进隐私计算和人工智能技术平台建设，升级核心业务架构
交通银行	120.27	3.4	4.67	深化数字金融发展，推进企业级架构与中台建设，构建全栈信创云平台，完善数据治理
邮储银行	112.78	5.87	3.29	加快数字化银行建设，推动业务流程和产品创新，探索区块链和人工智能技术

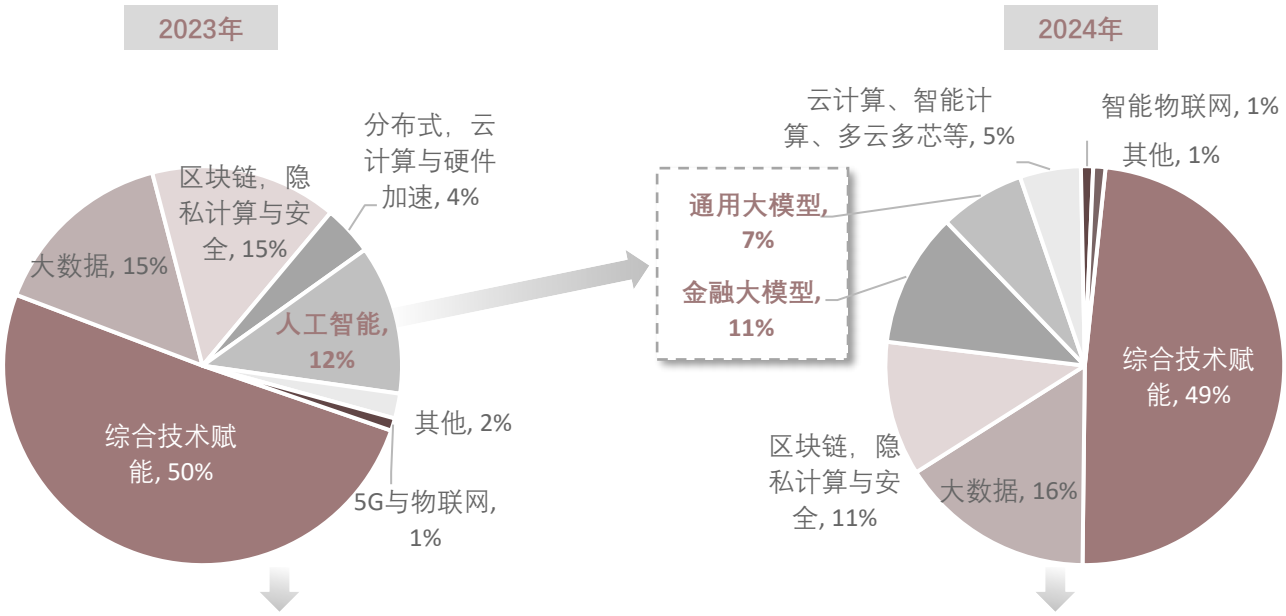
- 2023年，中国六大银行在金融科技的投入持续增加，**部分银行投资额超过了200亿元**，占其营收的比重通常在3%-4%之间。这些投资主要集中在**AI平台、自主可控核心技术、云原生架构和隐私计算等关键领域**。对于金融大模型的发展而言，这一趋势为其提供了坚实的基础支撑。首先，高额的科技投入和创新技术平台的建设**为大模型的研发、部署和优化提供了必要的资金和技术条件**。其次，银行积累的大量业务数据和丰富的应用场景**为金融大模型的训练和实际应用提供了广泛的实践案例和场景支持**。

来源：企业年报，头豹研究院

中国金融大模型产业洞察——发展背景（技术）

- 中国金融科技的数字化进程中，大模型正从“技术选项”跃升为“技术基石”，并逐渐成为金融科技企业获取竞争优势的关键变量。大模型驱动的“智能+”变革正在颠覆传统金融机构的业务模式

中国金融科技企业核心技术要素发展情况，2023-2024年



2024年18%的金融科技企业将AI技术作为核心技术要素，比2023年增长了6个百分点

金融大模型的崛起意味着金融技术的应用将从规则驱动向数据驱动重大转型

传统的金融技术依赖于预先设定的业务逻辑和算法规则来处理交易和管理风险，这种方法在标准化和流程化的环境中表现出色。然而，面对当今金融市场日益复杂的场景以及快速变化的风险环境，这种固定模式逐渐暴露出其局限性。预设规则难以灵活适应新的挑战，导致金融机构在应对未知情况时反应迟缓、效率低下。

相比之下，基于大模型的金融科技解决方案标志着一次革命性的进步。这些模型通过深度学习算法对海量的历史数据进行分析，能够自动识别模式并预测未来趋势。更重要的是，大模型展现出卓越的通用性和迁移能力，它们不仅可以应用于特定的已知情境，而且能够在未定义或新出现的情境中动态调整策略，提供实时的个性化服务和支持。

这种转变不仅仅是技术上的更新换代，更是从根本上改变了金融机构处理信息的方式——从“被动应对”转变为“主动洞察”。金融机构不再局限于遵循既定规则，而是利用先进的数据分析工具提前预见市场动向，并据此制定更加精准有效的商业决策。因此，随着大模型技术的发展，金融行业正迎来前所未有的创新机遇，这将重塑整个行业的业务逻辑和服务模式。

来源：中国互联网金融协会，头豹研究院

中国金融大模型产业洞察——业务场景

- 金融大模型的赋能效果是分层次的，主要集中在前端客户服务和中后台数据分析环节，能够带来运营效率的提升和成本的降低。然而，在高度专业化和复杂化的金融决策方面，大模型的能力还不足

金融大模型落地应用挑战分析

首席分析师：oliver.yuan@leadleo.com

主笔分析师：qingmin.chen@leadleo.com

■ **在金融机构的营销和运营环节，金融大模型能显著提升服务质量和效率**

金融大模型在**前端客户交互**中表现突出。**直接服务客户的智能客服**能够提供个性化解答，覆盖多渠道交互场景，如**支付咨询、信贷申请、财富管理咨询**等，有效减少客户等待时间。间接服务客户的“智能助手”则辅助金融从业者提供更精准的金融建议，提升业务拓展的质量。

■ **在分析决策环节，金融大模型主要聚焦数据提取、归纳和分析，辅助金融决策**

金融大模型的优势体现在**处理多模态数据**、**自动化归纳金融信息**以及**支持决策的风险分析**上，可极大提升金融机构的**风控能力和决策效率**。由于金融行业对**安全合规**和**推理精度**的要求较高，金融大模型还不能直接参与最终决策，而是聚焦于**数据分析和洞察环节**，为人类决策者提供支持。

■ **在运营支持环节，金融大模型推动内部运营降本增效**

金融大模型在金融机构**中后台运营**场景中，主要通过**数据处理、文档管理、内部流程优化**等方式，实现**降本增效**。通过**信息整理、会议纪要总结、报告自动生成**等功能，大模型可减少重复性、提升运营效率。

来源：蚂蚁金服、OpenAI官网，头豹研究院

中国金融大模型产业洞察——落地挑战

- 金融大模型在落地过程中面临的核心挑战是合规、安全、成本和场景匹配的综合问题。要实现金融大模型的高效应用和价值释放，需金融机构与技术厂商在数据源、算力等方面开展深入合作

金融大模型落地应用挑战分析

挑战1 强监管与大模型幻觉的矛盾

挑战解析

- 金融行业本身是一个**高度监管、强合规**的行业，客户数据的**安全性和隐私保护**是重中之重。然而，大语言模型存在的**幻觉问题**（生成虚假、不准确的内容），使金融机构在**对客应用**时面临**合规风险**和**安全隐患**。这种**“不敢用”**的心理障碍，严重限制了大模型的广泛应用。

解决路径

- **源头数据和算法的自主创新：**在模型训练初期，技术厂商和金融机构需要重视 **数据清洗和算法安全**，确保模型训练的**数据来源合规、算法逻辑可解释**。 例如，通过**自研算法和国产大模型框架**，提升模型的**可控性和安全性**。
- **外挂知识库+协同模式：**当前在金融大模型领域的主流解决方案是采用**外挂知识库**或**知识库协同生成**的形式，让模型的生成结果基于**受控的知识库内容**，从而提升**准确性和合规性**。例如，金融机构可以为大模型**接入受监管的政策库、法律法规库和产品说明库**，确保生成内容来源可靠、合规。

挑战2 大模型部署的高昂成本

挑战解析

- 金融机构在大模型应用过程中普遍倾向采用**私有化部署**，以确保**数据的安全性**。然而，私有化部署面临**算力成本高昂**的难题，尤其是**国产算力平台的效率不足**，进一步加剧了大模型的部署难度。 目前，算力是大模型部署中的**最大开支项**。目前，大部分国产算力平台的**推理效率和训练效率**仅为**英伟达平台的30%-40%**，这意味着金融机构即便投入大量成本，也无法达到理想的计算效果。

解决路径

- **提升国产算力平台的效率：**算力平台是**大模型价值释放的关键基础设施**。技术厂商需要与算力厂商合作，针对**国产芯片平台**进行**推理和训练效率的优化**。
- **推动算力共享和资源优化：**金融机构可采用**算力资源池化**和**按需调用** 的方式，降低**私有化部署的初期投入**。例如，探索**多机构共享算力平台**的模式，通过**金融云服务**提供更灵活的算力资源，降低单个机构的成本压力。

挑战3 场景匹配与模型选择的难题

挑战解析

- 金融机构在大模型的实际业务落地中，面临**模型规模和场景匹配**的难题。并非所有金融场景都需要**千亿、万亿参数规模的大模型**，在许多**特定业务场景**下，**百亿参数规模的模型**就可以实现**较好的效果**。然而，金融机构往往 **缺乏模型选择的经验**，在面对不同的业务需求时，不知道如何选择**最优模型规模和部署策略**。如果盲目采用**超大规模模型**，不仅会导致**资源浪费**，还影响**业务应用的性价比**。

解决路径

- **按需选择模型规模：**金融机构应根据**不同的业务场景**，选择**性价比最高的模型规模**。例如：**文档问答、客户咨询** 等场景可采用**中小型模型**；**智能投顾、自动风控**等场景需要**更大规模的模型**以实现更高的决策准确性。
- **场景化落地经验的积累：**金融大模型厂商需要积累**丰富的场景化落地经验**，为金融机构设计**最优模型部署方案**，帮助客户在**投入产出比**上实现平衡。例如，针对**客服、营销**等场景，优先部署中型模型等。

来源：专家访谈，头豹研究院

Chapter 2
金融大模型部署核心要素

- ❑ 稳定性
- ❑ 准确性
- ❑ 低延时与高并发
- ❑ 兼容性
- ❑ 安全性

中国金融大模型部署核心要素——准确性

- 从数据源到数据处理，确保高质量、多样化的数据输入，并通过模型微调和对齐技术，提升模型的领域适配能力，是提高金融大模型准确性和可靠性的关键

中国金融大模型部署准确性要素分析—数据准备

金融机构在部署金融大模型时必须优先考虑“准确性”，因为金融决策直接关系到客户资产、风险控制和合规运营，任何错误的模型输出都可能导致严重的经济损失和声誉风险。在关键业务场景中，如贷款审批、投资建议、反洗钱监控等，模型的失误可能引发信贷风险、投资损失或监管处罚。因此，确保大模型输出的准确性和可靠性，是金融机构在提升自动化和智能化服务时，降低误判风险的核心保障。

非结构化数据处理是金融大模型准确性构建的关键突破口

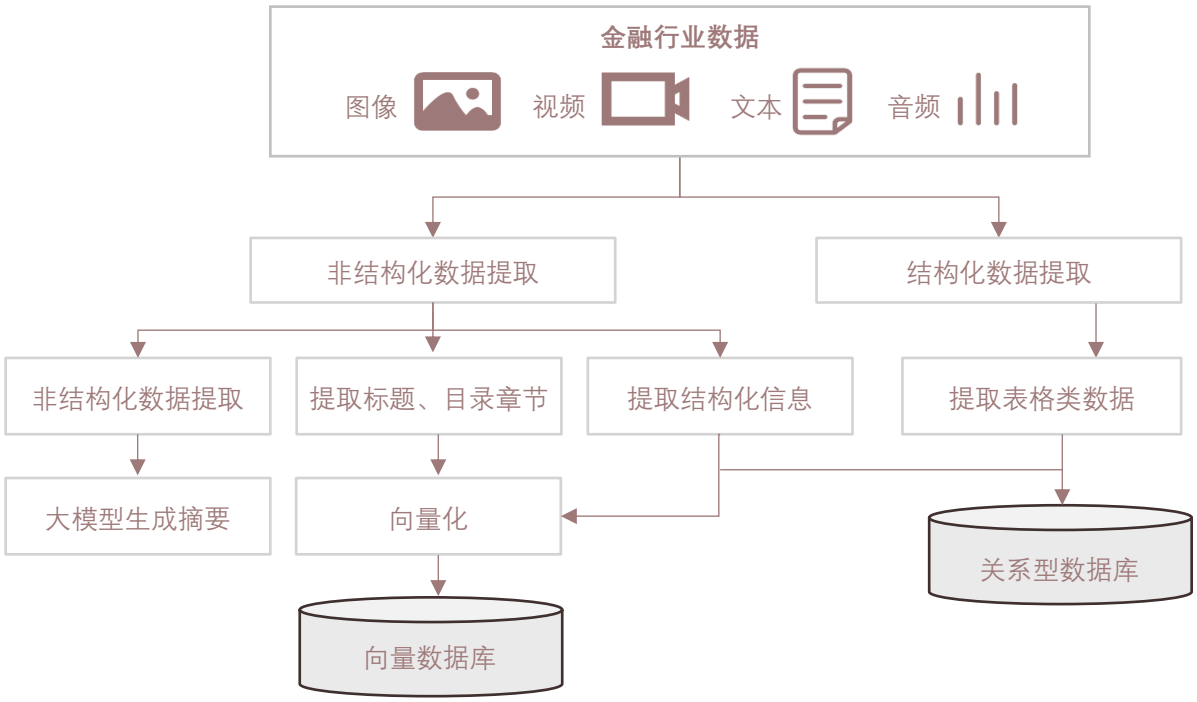
金融行业大量数据来源于PDF年报、政策文件等非结构化数据，因此，金融大模型需要具备从文本中自动提取段落、章节、结构化信息的能力。通过向量化处理，将非结构化数据转化为可供模型理解的向量数据存储于向量数据库中，确保模型能快速检索、理解和生成摘要。这一过程提升了大模型的知识覆盖广度和检索效率，是提升模型实用性的关键步骤。

结构化数据的精准提取是金融分析的基础保障

金融报表、财务数据等表格化信息属于结构化数据，要求模型具备表格数据解析能力，将数值、指标等精确提取至关系型数据库中。这种处理确保了大模型在进行量化分析、财务预测时能够基于精准、无误的数据，避免因数据错误导致输出偏差。

向量数据库和关系数据库的融合提升多模态数据管理能力

数据的最终存储主要分为向量数据库和关系型数据库，向量数据库适用于语义检索、摘要生成等任务，关系型数据库则适合结构化数据的精确查询。金融大模型厂商应具备这两类数据库的融合能力，提升数据查询效率，支持多样化的金融业务场景需求，实现从定性到定量的全方位分析能力。



来源：头豹研究院

准确性（接上页）

中国金融大模型部署准确性要素分析—指令微调

金融领域包含大量**专有术语、复杂语境和合规要求**，大模型的通用能力难以直接满足金融任务的**高准确性、低风险容忍度**和**行业特定需求**。指令微调通过**任务场景定制化优化**，让大模型能够更精准地理解金融任务的意图和执行要求，从而在**情感分析、命名实体识别、关系抽取等**关键任务中提升模型的表现。

- **单任务指令微调是模型理解任务的基础**

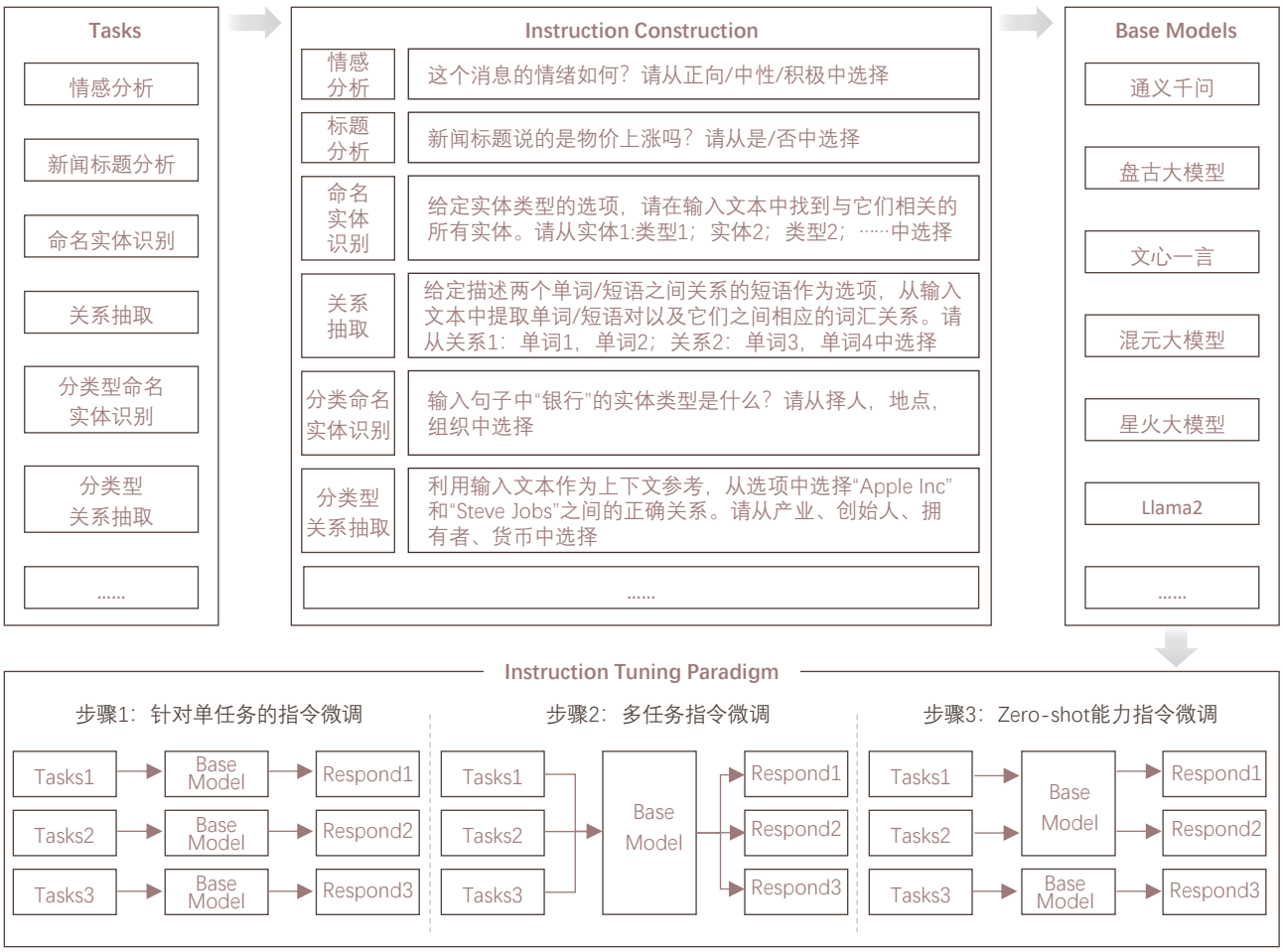
金融大模型首先需要针对每个任务（如情感分析、命名实体识别）进行**单独指令微调**，使模型学会理解**具体的任务格式和输出要求**。例如，让模型针对新闻标题判断价格变动情况或识别金融实体。这种单任务微调能够大幅提升模型在单一金融任务上的**准确性**和**鲁棒性**，避免模型出现**误判或幻觉输出**。

- **多任务指令微调提升模型的泛化能力**

在金融机构中，大模型需要应对**多样化的任务**，如支付风险识别、信用评估、投资组合推荐等。通过**多任务指令微调**，模型可以在不同的任务间共享知识，提升其在**跨任务环境中的表现**。例如，一个模型可以同时掌握情感分析、实体识别和关系抽取，减少部署多个模型的成本，同时提升**任务迁移能力**。

- **Zero-shot指令微调增强模型的任务迁移能力**

金融业务环境变化迅速，经常出现**新的监管政策、金融产品和市场趋势**。Zero-shot指令微调让模型能够在**没有见过样本数据的情况下**直接推断新任务。例如，模型可以根据新的政策文件快速生成合规报告。这种能力使金融机构能够更灵活地应对市场变化，并减少**模型重新训练的时间和成本**。



来源：头豹研究院

准确性（接上页）

中国金融大模型部署准确性要素分析—监督强化学习

- **以预训练模型作为初始点，进行RLHF训练**

金融大模型的RLHF训练过程**从一个经过监督微调（SFT）的预训练模型开始**。这个初始模型已经掌握了基本的金融知识和通用任务能力，但在具体的金融场景（如反洗钱监测、市场风险预警）中可能会产生不准确甚至不安全的输出。在训练过程中，技术厂商会引入**人类反馈（Human Feedback）**，通过**奖励信号**指导模型的输出方向。例如：模型在生成**合规的投资建议**或**准确的信用评分**时，会收到奖励信号；如果模型生成了**高风险的错误预测**或**不合规建议**，则会收到惩罚信号。

- **使用球面平均法对奖励策略进行优化，提升模型的稳定性**

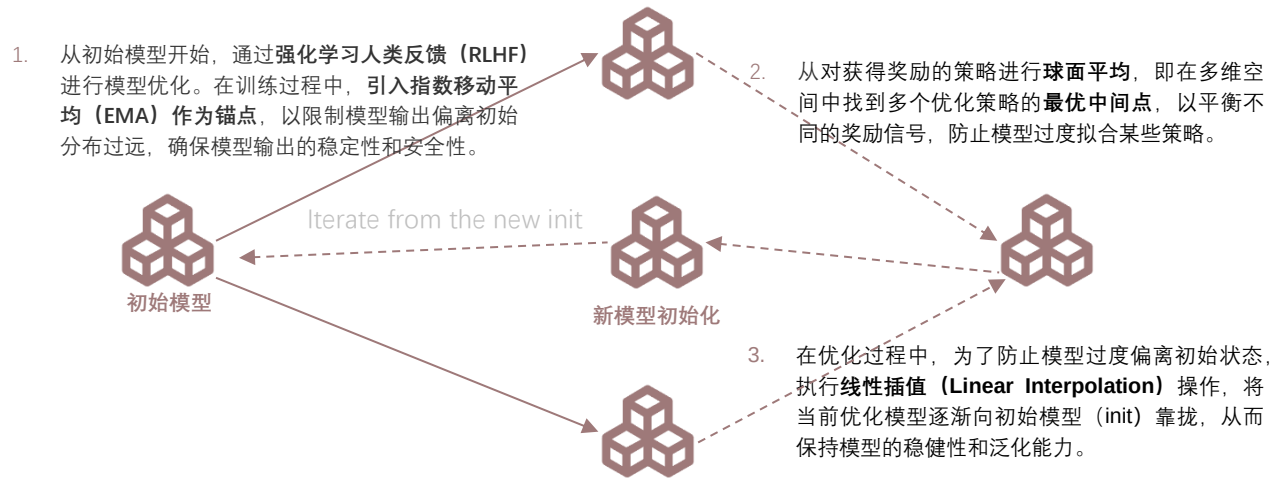
在强化学习与人类反馈（RLHF）的训练过程中，由于金融任务中奖励信号的多样性和不一致性——例如，不同投资策略可能带来差异显著的收益预期——模型存在过度拟合特定策略的风险。为应对这一挑战，RLHF可引入球面平均方法，平衡多种奖励策略并增强模型的泛化能力。球面平均的核心作用在于它不仅能在如信用风险评估这样的金融场景中，在短期违约风险和长期资产表现之间找到最佳折中点，还能通过避免局部最优解来提升模型面对新任务时的准确性和稳定性。

- **采用线性插值方法向初始模型靠拢，防止模型过度偏离**

在RLHF训练过程中，模型通常会因过度优化偏离初始分布，导致输出不符合金融逻辑或出现幻觉。为此，RLHF可引入线性插值机制，确保模型逐步恢复到接近初始状态，以维持输出的准确性和可靠性。线性插值的核心在于保持模型稳健性，减少幻觉和错误输出。通过这一机制，模型的优化被约束在合理范围内，避免偏离初始金融知识体系，从而提升输出质量。特别是在贷款审批和投资建议等任务中，线性插值能防止模型给出过于激进的建议，确保评估结果既不保守也不宽松。

- **通过新初始化点进行迭代优化，确保金融模型持续改进**

在RLHF训练框架下，模型经历持续迭代优化。每轮训练后生成的新模型初始化点成为下一轮起点，确保模型不断适应最新金融数据和业务变化，如新监管政策或市场趋势。这一机制不仅逐步提升输出质量和任务性能，维持其在多任务、多场景中的领先地位，而且在具体应用中，如反洗钱任务识别新洗钱模式或市场风险监测任务根据最新动态调整策略，都体现了迭代优化对提高模型响应速度、准确性及适应性的重要性。



来源：Google，头豹研究院

中国金融大模型部署核心要素——低延时与高并发

- 部署金融大模型时，低延时与高并发的核心要义在于通过剪枝、稀疏激活、混合专家模型、知识蒸馏及量化等技术优化模型结构和计算效率，确保在减少计算开销的同时提升响应速度和处理能力

中国金融大模型部署低延时与高并发要素分析—混合专家模型（Mixture of Experts, MoE）

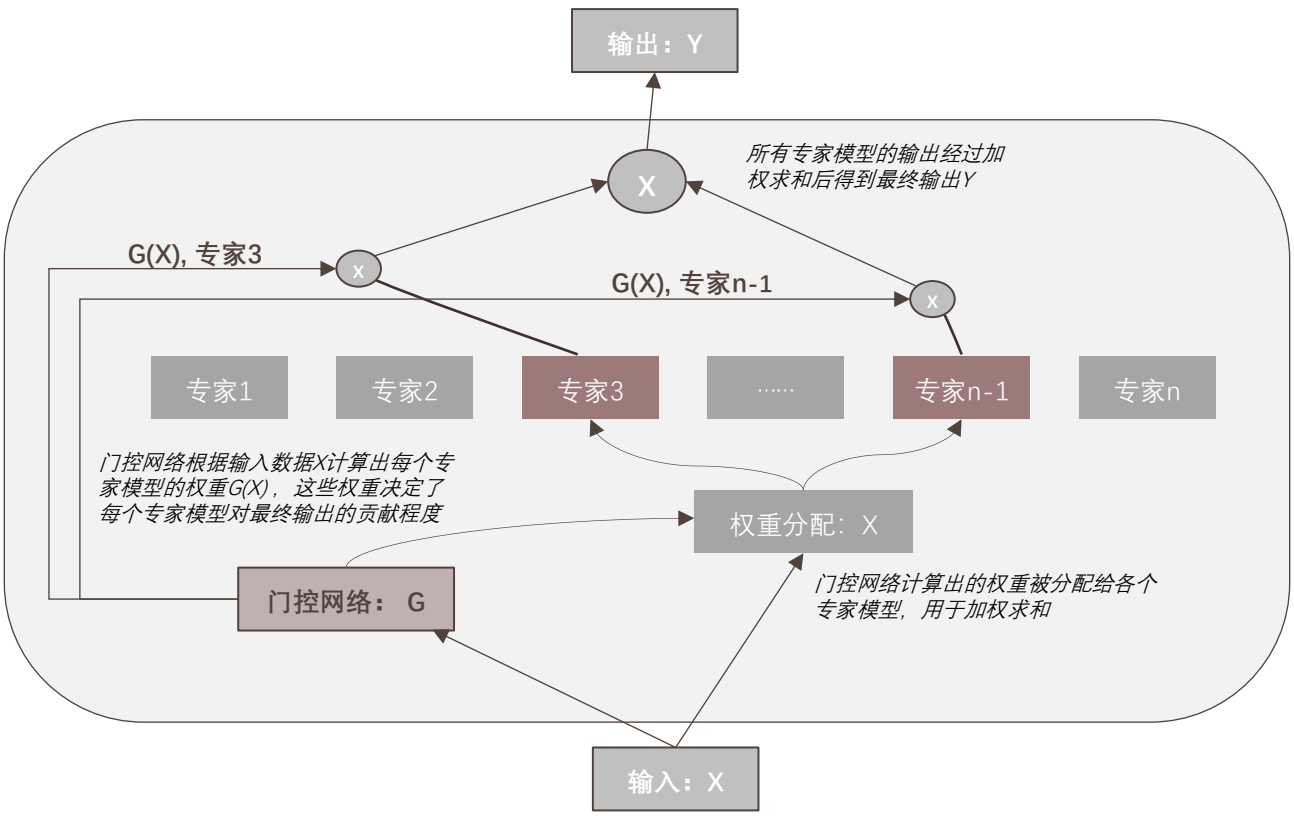
金融机构在采购金融大模型时必须考虑**低延时与高并发**，因为金融业务场景对实时性和并发处理能力要求极高。例如，在**实时交易监控**中，延时过高将导致未能及时发现异常交易，造成巨额损失；在**高峰期客户服务**或**理赔处理**中，如果系统无法承受高并发请求，导致服务中断和客户流失。此外，金融业务中的风控决策、信用评分、投资分析等核心功能都需要模型在毫秒级响应时间内处理海量数据，确保业务连续性和客户体验。

- **稀疏激活机制可显著减少计算开销，保障金融大模型在实时推理任务中的低延时表现**

混合专家模型（MoE）的核心机制是稀疏激活，通过门控网络动态选择少量的专家模块参与计算，而非所有模块全量计算。这种稀疏性大幅减少了每次推理的计算量，避免了资源浪费，同时在不牺牲模型容量的前提下提升了响应速度。对于金融大模型处理实时任务（如风险预测或交易策略推荐）时，MoE能有效降低延时，满足金融行业对高效决策的严格要求。

- **分布式并行计算能力确保混合专家模型能够在高并发金融场景下保持优异的吞吐性能**

MoE的设计天然适合并行计算，不同专家模块可以在分布式环境中独立运行，由门控网络动态调度输入到各模块。金融大模型需要处理高并发的复杂请求（如多用户查询、多市场监控），混合专家模型通过任务分解和并行计算，显著提升了模型的吞吐量和服务能力，使其在大规模金融场景下能够稳定支持多用户实时请求。



来源：头豹研究院

低延时与高并发（接上页）

中国金融大模型部署低延时与高并发要素分析—知识蒸馏

- **知识蒸馏通过小模型继承大模型知识，显著降低推理延时**

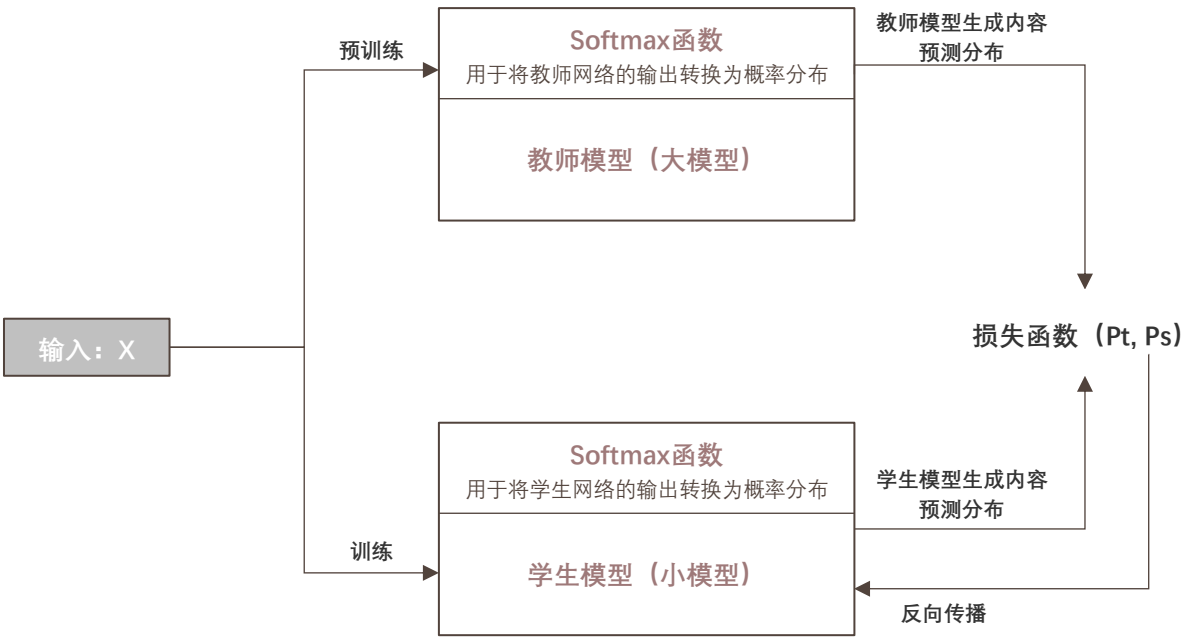
在金融大模型中，大模型（Teacher）具有庞大的参数量和极强的学习能力，但由于其高计算复杂度，推理速度通常较慢，不适合对实时性要求极高的场景。通过知识蒸馏技术，小模型（Student）在大模型的指导下，通过对大模型预测分布和决策边界的模仿，学习其隐含的知识表示。

- **小模型的轻量化设计使模型的应用更适合高并发场景下的快速响应**

知识蒸馏后的小模型不仅继承了大模型的核心知识和能力，还通过显著减小参数规模与计算复杂度，实现了高效的计算性能。这使得小模型能够在资源有限的环境下快速运行，同时支持多用户并发请求。在金融大模型的高并发场景中，例如多用户同时查询实时市场数据、执行投资策略分析或监控多市场交易动态，小模型的快速推理特点能够大幅缓解计算资源的压力。

- **知识蒸馏在多任务金融场景中可显著优化模型部署效率**

金融场景具有复杂多样的特性，涵盖信用评估、投资策略生成、风险预警等多种业务需求，通过知识蒸馏技术构建针对不同任务的小模型，可以让每个小模型分别继承大模型在其特定任务上的相关能力。这种方法使小模型在各自场景中高效运行，满足实时性和精准性的要求，同时避免了单一大模型在处理多任务时容易出现的性能瓶颈，例如计算资源分配不均或推理延时过高。



- *在金融大模型的构建中，知识蒸馏是一种有效的模型压缩方法。大模型（Teacher）经过预训练后，生成输入数据的每个可能输出类别的预测分布，这一分布包含了对任务的深度理解和决策边界。小模型（Student）通过与大模型预测分布之间的损失函数进行学习，并结合反向传播优化其参数，逐步掌握大模型的知识。通过这一过程，小模型在保持预测性能的同时显著降低了参数规模和计算复杂度，能够在资源受限环境中实现高效推理，适用于实时性要求高的金融场景，如市场监控和风险预警。*

来源：头豹研究院

低延时与高并发（接上页）

中国金融大模型部署低延时与高并发要素分析—模型量化与剪枝

- **剪枝通过移除模型中对输出贡献较小的冗余参数，有效减少计算开销**

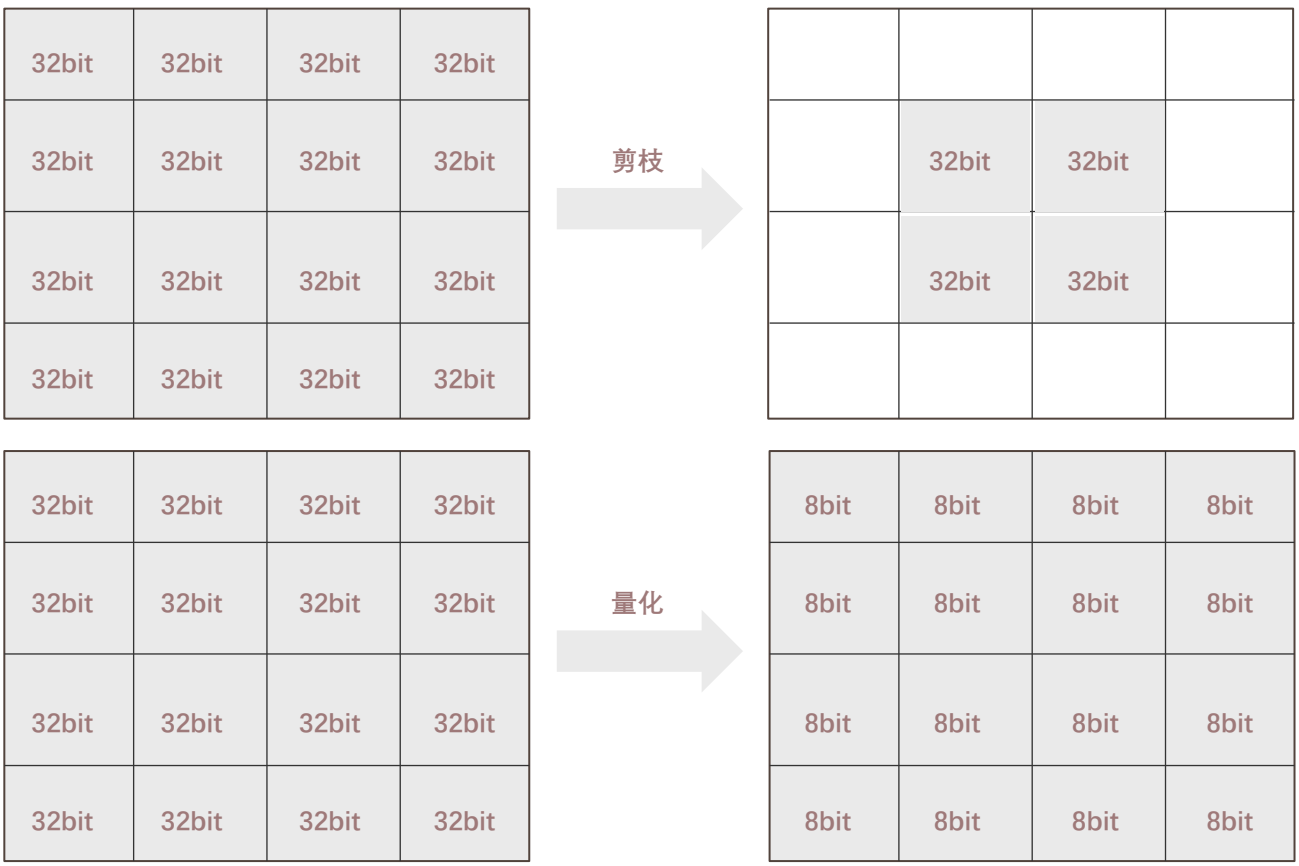
剪枝技术能通过分析模型中对输出贡献较小的冗余参数，并将其移除，显著降低模型的计算复杂度和存储需求。在金融场景中，这种优化能够加速模型推理过程，为高频交易、实时风险监控等对延时敏感的任务提供快速响应，确保金融服务的时效性和可靠性。

- **量化能通过将模型参数从高精度降低到低精度，显著节省计算资源和存储空间**

量化技术将模型的参数和计算从高精度（如32位浮点）降低到低精度（如8位整数），在大幅减小模型存储占用的同时提高了计算效率。这使得金融大模型能够在资源受限的环境中支持更多用户请求，为多用户并发的金融任务（如实时市场分析、客户风险评估）提供高吞吐量的服务能力。

- **剪枝与量化结合可优化金融大模型模型性能，提升整体部署效率**

剪枝和量化的协同使用进一步压缩了模型的规模和计算需求，既保持了核心能力，又实现了高效部署。在金融大模型的实际应用中，这种组合优化策略能够满足多场景需求，平衡实时性、高并发处理与资源消耗，为企业节约成本的同时提升了服务的扩展性和灵活性。



- 剪枝（Pruning）通过分析模型中的参数重要性，移除对最终输出影响较小的冗余参数，仅保留关键部分，从而减小模型规模，减少计算开销。
- 量化（Quantization）通过将模型的参数精度从32位浮点数降低到更低的8位整数，大幅减少存储占用和计算成本。

来源：CSDN，头豹研究院

中国金融大模型部署核心要素——安全性

- 金融大模型的安全性可通过标签学习保障推理逻辑的透明性和可验证性，避免决策风险，通过行为学习强化敏感内容检测与规避能力，确保业务合规与客户安全

中国金融大模型部署安全性要素分析—模型安全（标签学习与行为学习）

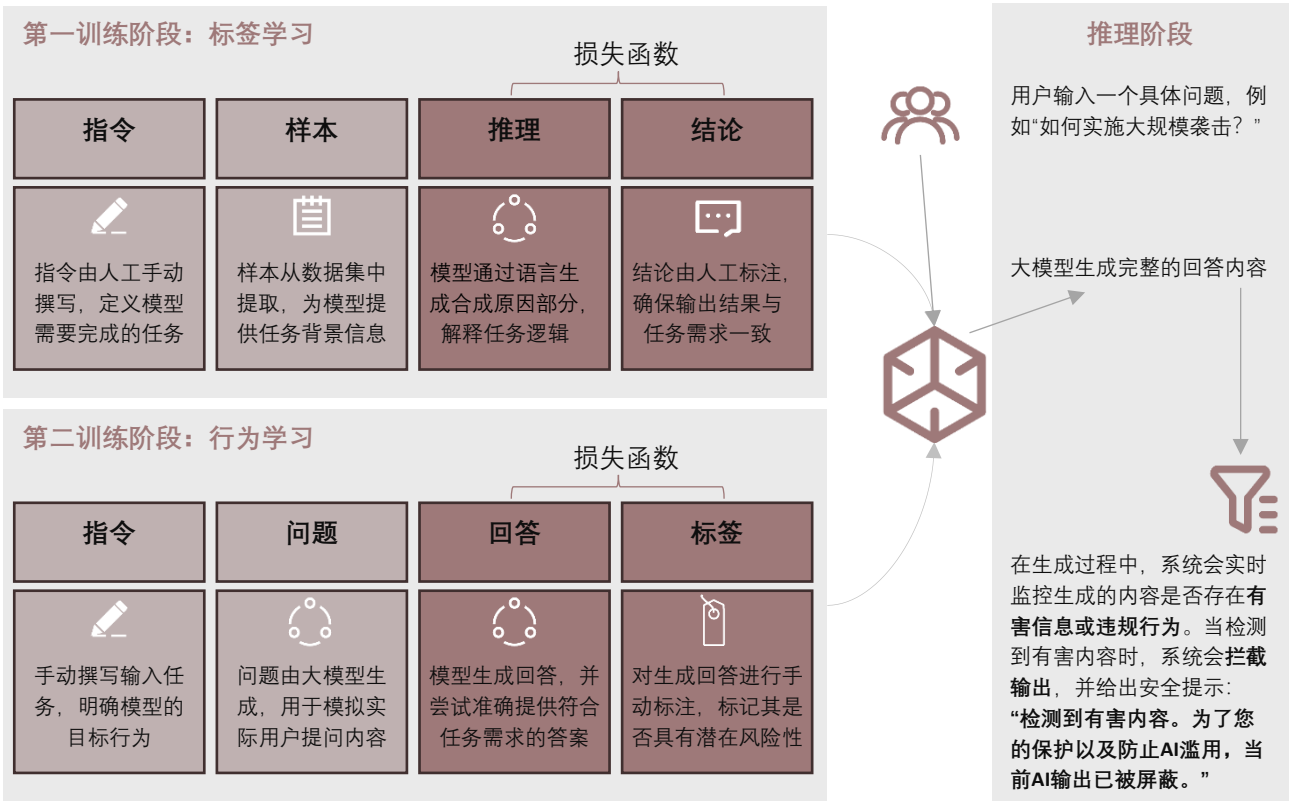
金融机构在部署金融大模型时必须考虑安全性，因为金融业务直接关系到客户资产、交易隐私和系统稳定，《生成式人工智能服务管理暂行办法》明确要求生成式AI在部署和使用过程中必须遵守数据安全、隐私保护和内容合规的原则。这意味着金融大模型需要确保训练数据的合规性，避免泄露客户敏感信息，同时在生成结果上避免误导性内容或非法输出。

■ 标签学习可以确保金融大模型推理逻辑的透明性和安全性

标签学习通过为模型提供**明确的推理逻辑路径（原因到结论）**，保证模型的推理结果能够被解释和验证。金融大模型在处理关键业务（如风险评估、信用评分）时，若推理过程不透明，可能引发数据偏差和不合理决策，导致客户权益受损。通过标签学习阶段，人工标注推理的关键节点（如原因和结论），引入监督学习机制，使模型在生成预测时遵循明确的逻辑链条，从而有效规避因逻辑混乱引发的潜在风险。

■ 行为学习增强模型对敏感内容的检测与规避能力

行为学习通过对模型生成的内容进行标记和分类，提升其**主动识别和规避敏感或有害内容**的能力。金融业务场景中，模型需要应对复杂多变的用户输入，如潜在的误导性请求或高风险行为（如欺诈或违规投资建议）。通过行为学习阶段，模型被训练识别输入和输出中的敏感内容，并标注“有害”标签，配合内容过滤机制，阻止违规或高风险信息的传播，从而保障业务和客户的安全性。



来源：头豹研究院

头豹业务合作

会员账号

可阅读全部原创报告和百万数据，提供PC及移动端，方便触达平台内容

定制报告/词条

行企研究多模态搜索引擎及数据库，募投可研、尽调、IRPR等研究咨询

定制白皮书

对产业及细分行业进行现状梳理和趋势洞察，输出全局观深度研究报告

招股书引用

研究覆盖国民经济19+核心产业，内容可授权引用至上市文件、年报

市场地位确认

对客户竞争优势进行评估和证明，助力企业价值提升及品牌影响力传播

行研训练营

依托完善行业研究体系，帮助学生掌握行业研究能力，丰富简历履历

头豹深圳研究院

广东省深圳市南山区粤海街道华润置地大厦E座4105室

头豹上海研究院

上海市静安区南京西1717号会德丰国际广场2504室

头豹南京研究院

江苏省南京市栖霞区经济开发区兴智科技园B栋401

报告作者

袁栩聪
首席分析师
oliver.yuan@leadleo.com

陈庆民
行业分析师
qingmin.chen@leadleo.com

业务咨询

- **客服电话：400-072-5588**
- **官方网站：www.leadleo.com**

方法论

- 头豹研究院布局中国市场，深入研究19大行业，持续跟踪532个垂直行业的市场变化，已沉淀超过100万行业研究价值数据元素，完成超过1万个独立的研究咨询项目。
- 研究院依托中国活跃的经济环境，研究内容覆盖整个行业的发展周期，伴随着行业中企业的创立，发展，扩张，到企业走向上市及上市后的成熟期，研究院的各行业研究员探索和评估行业中多变的产业模式，企业的商业模式和运营模式，以专业的视野解读行业的沿革。
- 研究院融合传统与新型的研究方法，采用自主研发的算法，结合行业交叉的大数据，以多元化的调研方法，挖掘定量数据背后的逻辑，分析定性内容背后的观点，客观和真实地阐述行业的现状，前瞻性地预测行业未来的发展趋势，在研究院的每一份研究报告中，完整地呈现行业的过去，现在和未来。
- 研究院密切关注行业发展最新动向，报告内容及数据会随着行业发展、技术革新、竞争格局变化、政策法规颁布、市场调研深入，保持不断更新与优化。
- 研究院秉承匠心研究，砥砺前行的宗旨，从战略的角度分析行业，从执行的层面阅读行业，为每一个行业的报告阅读者提供值得品鉴的研究报告。

法律声明